# Potential Implications of Changes in ChalleNGe Admission Criteria: A Literature Review

Lauren D. Malone and Jennifer R. Atkin

April 2016

CNA
ANALYSIS & SOLUTIONS

This document contains the best opinion of CNA at the time of issue.
It does not necessarily represent the opinion of the sponsor.

**Distribution**

Distribution unlimited. Specific authority: N00014-16-D-5003.

Copies of this document can be obtained through the Defense Technical Information Center at www.dtic.mil or contact CNA Document Control and Distribution Section at 703-824-2123.

**Photography Credit:** Louisiana National Guard Youth Challenge Program. Photo by Megan Ready. http://www.langycp.com/news-media/news/ycp-press-release-22-215-cadets-graduate-from-camp-minden

**Approved by:** **April 2016**

Jeffery M. Peterson, Director
Fleet and Operational Manpower Team
Resource Analysis Division

# Abstract

In this report, we consider two potential changes to the admission criteria of the National Guard Youth ChalleNGe Program (ChalleNGe)—limiting the ages for admission and introducing a standardized test for admission—both with the aim of maximizing cadets' growth at ChalleNGe. Restricting the eligible ages could optimize the potential for noncognitive growth; a minimum admission score could maximize cognitive improvement. We synthesize the literature in these areas and ultimately determine that neither change is recommendable. There *are* age-related variables that affect noncognitive development, making it less likely to occur at younger ages and thus more likely to be significantly improved at ChalleNGe. However, we do not recommend excluding older, at-risk youth from the program based solely on the desire to achieve maximum noncognitive growth. In addition, a standardized test score is insufficiently accurate as a representation of true ability to be used as an admission criterion.

This page intentionally left blank.

# Executive Summary

The National Guard Youth Challenge Program (ChalleNGe) is a quasi-military, 22-week residential program designed to serve 16- to 18-year-old high school dropouts as well as students at risk of dropping out (students who have earned fewer credits than expected are considered at risk of dropping out). Participants are referred to as cadets. They live in barracks, wear military-style uniforms, and perform activities typically associated with military training (e.g., marching, drills, and physical training). Participation, however, is voluntary and cadets have no subsequent requirement for military service. The ChalleNGe model has both academic and noncognitive components, such as leadership/followership, responsible citizenship, service to community, life-coping skills, physical fitness, health and hygiene, job skills, and academic excellence.

ChalleNGe leadership has recently focused on finding ways to maximize cadets' growth over the course of the program—in terms of both their cognitive and noncognitive skills. Its current focus lies in determining whether further restrictions to admission criteria should be considered to maximize such growth. ChalleNGe leadership has asked CNA to evaluate, first, whether cadets' age ranges should be further restricted (due to an increased likelihood of noncognitive growth at certain ages) and, second, whether there is an optimal range of incoming test scores for maximizing cognitive growth (and thus if ChalleNGe should impose a test-score admission criterion). Our future work will quantitatively evaluate these questions, using data collected in support of our seven-site cadet survey. In this report, we review the relevant literature. Our synthesized findings suggest that neither of the proposed policy changes is recommendable.

## Noncognitive growth and age

In considering whether ChalleNGe should further restrict the age range of incoming cadets (currently 16-18) in order to maximize noncognitive growth, the following themes in the literature should be taken into account:

- No studies find a *direct*, independent impact of age on noncognitive growth. That said, noncognitive growth *is* correlated with a number of age-dependent

variables, such as school-grade progression (noncognitive development is more likely in higher grades) and previous experience.

- Most noncognitive growth occurs *outside* the 16-18 age range. In addition, early investments are important, so the earlier adolescent interventions occur, the larger the impact they likely will have on noncognitive development.

Based on these findings, increasing the size of the 16-year-old population would likely maximize the potential for noncognitive growth—those arriving with less-well-developed noncognitive skills are those for whom more significant improvement is feasible—but such a restriction would leave many 17- and 18-year-old at-risk youth without access to the ChalleNGe program. This would unfairly prevent older youth from benefiting from the program and likely would limit the programs' placement rates since older cadets, in general, are more employable after completing ChalleNGe because of employers' age restrictions. We therefore do not recommend altering the age restrictions or placing greater emphasis on admitting younger cadets solely in an attempt to maximize the level of noncognitive growth.

# Test scores and potential cognitive growth

The literature's primary findings regarding the strength of test scores as a predictor of future success (and thus the extent to which they should be relied on as an admission criterion) include the following:

- Test scores are significantly less predictive of success (whether in terms of schooling, employment, or other outcomes) than other measures that also incorporate noncognitive skills.

- Test scores are riddled with a number of "pollutants"—factors that occlude the interpretive value of test scores as a measure of inherent cognitive ability. They include test anxiety, previous classroom and schooling experiences, student motivation, stereotype threat (in which the fear of conforming to a preexisting stereotype reduces people's capacity for focusing on the task at hand), and the practice of educators "teaching to the test."

Thus, although initial test scores may give an indication of feasible cognitive improvement, the information revealed by test scores is too muddled with other information to make them reliable indicators of cadets' true abilities. In addition, an initial test-score admission criterion would remove all those scoring below the cutoff—perhaps those most in need of the program's support. We therefore do not recommend that ChalleNGe restrict the eligible population based on test scores.

# Contents

This page intentionally left blank.

# List of Figures

This page intentionally left blank.

# Glossary

| | |
|---|---|
| ACT | American College Test |
| AFQT | Armed Forces Qualification Test |
| ChalleNGe | National Guard Youth ChalleNGe Program |
| GED | General Educational Development |
| GRE | Graduate Record Examination |
| HiSET | High School Equivalency Test |
| SAT | Scholastic Assessment Test |
| SES | Socioeconomic Status |
| TABE | Test of Adult Basic Education |

This page intentionally left blank.

# Introduction

The National Guard Youth ChalleNGe Program (ChalleNGe) is a program designed to provide a second chance to 16- to 18-year-old high school dropouts and support for those at risk of dropping out. The program consists of a 5-month residential phase and a 12-month nonresidential mentoring follow-up. ChalleNGe has a quasi-military structure: participants (known as cadets) live in barracks, wear military-style uniforms, and perform activities typically associated with military training (e.g., marching, drills, and physical training). However, cadets participate voluntarily and have no subsequent requirement for military service. ChalleNGe's primary objective is to help "young people improve their self-esteem, self-confidence, life skills, education levels, and employment potential" [1].

The individual ChalleNGe academies (there are 35 locations across the country) are given discretion in how they structure their programs and which academic options they offer. At some academies, cadets can earn a high school diploma, some programs provide credit recovery so that cadets can earn high school credits and return to high school, and sometimes the focus is on cadets passing the General Educational Development (GED) test or High School Equivalency Test (HiSET). A number of programs offer some combination of these three options, and still others are equivalent to high schools and award state-certified high school diplomas.

In addition to providing an academic program, ChalleNGe seeks to instill life skills in the cadets. Toward that end, the core values of ChalleNGe are honor, courage, and commitment. The program also has eight core components: leadership/followership, responsible citizenship, service to community, life-coping skills, physical fitness, health and hygiene, job skills, and academic excellence [2]. All of these core values and components help cadets focus on the changes needed to become productive citizens on completion of the ChalleNGe program.

Our previous evaluations of ChalleNGe's success have largely focused on the cadets' growth in both cognitive and noncognitive skills over the course of their time in the program ([3], [4]). There has been increased interest in recent years on noncognitive growth because of the importance of certain attributes for success in the labor market: self-discipline, timeliness, determination, respectful citizenship, and self-confidence. Increased employability is a primary ChalleNGe objective, so evaluating whether cadets experience improvements in these characteristics has been a reasonable way to determine whether their experiences at ChalleNGe actually

improve their employability. Thus, the ChalleNGe leadership seeks to determine ways to maximize cadets' noncognitive growth, and its current focus lies in whether there is an optimal age (or age range) for such growth to occur. That is, if the program is primarily interested in maximizing noncognitive growth, is there a particular age range when cadets are most likely to achieve this? And should the program therefore further limit the acceptable ages of ChalleNGe cadets?

The primary means of measuring cognitive improvement has been improvements in cadets' scores on the Test of Adult Basic Education (TABE). In an effort to maximize cognitive growth, ChalleNGe leadership recently asked whether there is a specific range of incoming TABE scores in which the potential for score growth is greatest. If such a score range can be identified, they might consider implementing a test-score cutoff for acceptance into the program; this could apply to all cadets or to some fraction of the incoming class. Although these questions lend themselves to data analysis—and we will evaluate both the impact of age on noncognitive growth and the impact of initial TABE scores on TABE growth in our cadet survey analysis—they can also be informed by the existing literature. In this report, we review the literature and make corresponding recommendations in the following areas:

- Is there a relationship between age and noncognitive skill development? Is there a particular age or age range at which noncognitive growth is most likely to occur or is more receptive to influence? Based on these findings, should ChalleNGe consider further restricting cadets' acceptable age range?

- Is there a particular range of pre-TABE scores for which the likelihood of achieving maximum TABE growth is highest? What has the literature found to be the strongest predictors of test scores and test score improvement? Why might test scores be "polluted" and thus not be as strong a measure of ability or achievement as they are often assumed to be? Based on these findings, should ChalleNGe implement a minimum required test score?

We did not limit our literature search to any particular academic areas, but the bulk of the literature we discuss comes from the fields of economics, education, and psychology since these are the fields that have most closely addressed our questions.

The remainder of the paper is organized as follows. In the next section, we review the literature on noncognitive skill growth. This includes a discussion of the general importance of noncognitive skills and the determinants of noncognitive skill development. We offer recommendations based on what we found in the literature. Then we review the literature's findings on maximizing test score gains. We also review what previous studies have found regarding the strength of test scores in predicting success, other factors captured by test scores (including teachers' effectiveness, socioeconomic status, and stereotype threat), and the possible effects of "teaching to the test" if test score improvements are overemphasized. Once again, we offer recommendations based on the literature. In the final section, we conclude.

# Noncognitive Skill Growth and Age: Is There a Relationship?

In this section, we review a number of studies that evaluate the determinants of noncognitive skill development. Although noncognitive growth can occur throughout one's life and starts at very early ages, we focus on adolescent and early-adult noncognitive development since these are the relevant age ranges for ChalleNGe. We begin by defining terms that will be relevant to our discussion of noncognitive skills. To motivate our discussion, we then summarize findings that noncognitive skills are important for educational and labor market success. That is followed by a review of the literature's findings regarding important determinants of noncognitive growth, including both those studies finding that non-age-related factors are most important and those indicating that age *is,* in fact, important. Synthesizing these findings, we are then able to make recommendations for ChalleNGe regarding whether—based on the literature—the program should consider further restricting the age range of incoming cadets.

## Noncognitive skills: Relevant terminology[1]

The term *noncognitive skills* can be thought of as referring to all skills that are not academic in nature. Thus, the ability to solve long-division problems is a cognitive skill, whereas conscientiousness, perseverance, leadership, and positive attitude are noncognitive skills. Because of their all-encompassing nature, it is difficult to measure noncognitive skills with precision. During the middle of the last century, psychologists focused on developing so-called social learning theories—theories that account for how humans behave in complex social situations. In particular, the concepts of *locus of control, self-efficacy,* and *self-concept* were developed to explain why different people, faced with the same incentives, make different decisions.

[1] Large portions of this section are taken directly from Wenger and McHugh 2008 [5].

### Locus of control

Locus of control describes the extent to which a person believes that rewards are contingent on or closely related to his or her behavior. Those who believe in a strong connection between their behavior and eventual outcomes have an *internal* locus of control; those who believe the connection is weak have an *external* locus of control. Those with an external locus of control may believe that outcomes are controlled by luck or fate, or by other powerful persons; in either case, the important distinction is the extent to which one's own behaviors are thought to affect outcomes. Essentially, the scale typically used to measure this evaluates the extent to which respondents believe that they can control their lives.

### Self-efficacy and self-concept

Individuals' beliefs about their ability to perform well at a particular task make up their self-efficacy. Because of its task specificity, it is often referred to by subject-specific terms, such as math efficacy or science efficacy. Whereas self-esteem focuses on judgements of self-worth, self-efficacy focuses on judgements of personal capabilities [6]. Specifically, Bandura (1986) defines perceived self-efficacy as "people's judgements of their capabilities to organize and execute courses of action required to attain designated types of performances. It is concerned not with the skills one has but with judgements of what one can do with whatever skills one possesses" ([7], p. 391). That is, self-efficacy defines a person's beliefs about whether he or she can complete the task at hand, given his or her skills.

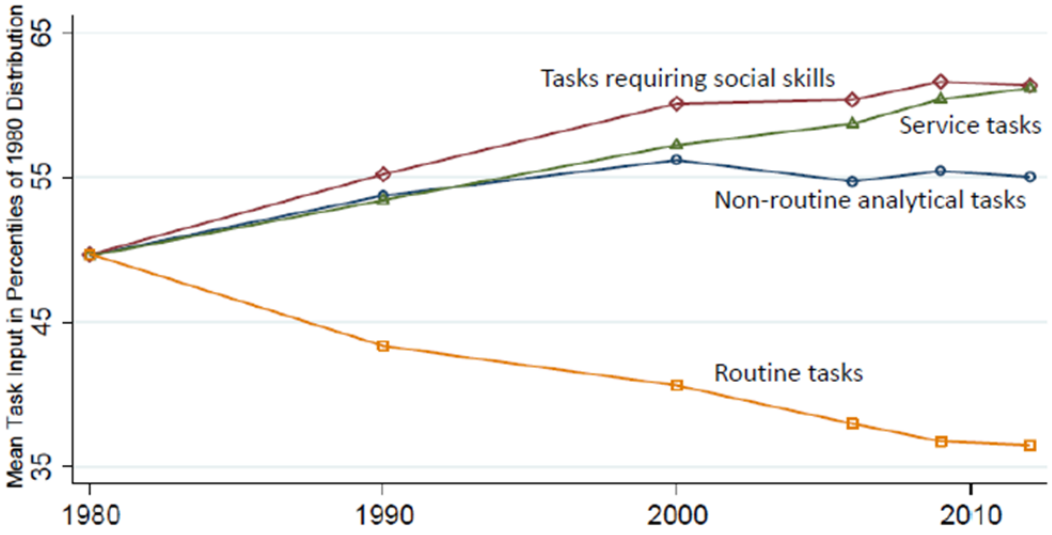
While self-efficacy beliefs focus on specific tasks or situations, self-concept is a more general measure. Specifically, "Self-concept…contains many perceptions about the self, including self-efficacy…[and] is developed as a result of external and internal comparisons, using other people or aspects of the self as frames of reference" ([6], p. 186). Conversely, self-efficacy focuses only on whether *I* can complete a certain task, regardless of how that compares with my beliefs about my peers' abilities [6].

## Noncognitive skill development matters

In recent decades, there has been increased attention—among scholars and employers alike—on the importance of noncognitive skills in personal development and people's ability to ultimately succeed in the labor market. A sizable set of research has shown that, in addition to cognitive skills and typical metrics of ability and achievement, prospective employers should also be paying attention to candidates' noncognitive skills (and they are).

In a recent discussion of the emergence of the computer and advanced technologies, Deming notes that employment in "routine occupations" has been falling since the 1980s [8]. In addition, the jobs that are higher skilled and harder to automate—jobs in which employees are more likely to find themselves as the automatable jobs disappear from the job market—are increasingly requiring more social skills [8]. Figure 1 illustrates these changes in the distribution of worker tasks. In this figure, each task type has a mean of 50 percentiles in 1980, by construction. Thus, changes since 1980 should be interpreted as "changes in the employment-weighted mean of each task relative to its importance in 1980" ([8], p. 9). The prevalence of tasks requiring social skills grew by 24 percent from 1980 through 2012, whereas so-called non-routine analytical tasks grew by only 11 percent over the same period [8].

Figure 1. Worker tasks in the U.S. economy[a]




Source: [8].

[a.] Based on 1980-2000 Censuses and 2005-2013 American Community Surveys.

It is not surprising that a number of studies have shown that noncognitive skills are important both for development and for a variety of success metrics. Levin, for example, finds that the relationship between test scores and later earnings is small, whereas there is a larger link between educational attainment and economic outcomes [9]. He suggests that this is due to the noncognitive aspects of education, which cannot be captured by standardized test scores. Specifically, he notes that education has effects on "the development of interpersonal and intrapersonal skill and capabilities that affect the quality and productivity of the labour force" ([9], p. 269). In order to simultaneously meet the economic, political, and social demands of the workforce, more is required of employees than the cognitive proficiencies

demonstrated by test scores. Both interpersonal (e.g., the ability to relate to others) and intrapersonal (e.g. good judgement, self-sufficiency) skills are required to be effective in the workforce [9]. It therefore comes as no surprise that employers indicate on surveys that the most important employee attributes are self-discipline, punctuality, attendance, goal setting, taking responsibility, and listening skill; academic abilities place much lower on the list than all of these attributes [9]. Similarly, in the classroom setting, students with higher self-efficacy have been found to be more productive learners. Research shows that those who have confidence in their work monitor their time more effectively, are more efficient problem solvers, are more persistent, work harder, and evaluate their progress more frequently than their equally able peers [10].

Among the most well-known research establishing the importance of noncognitive skills is that by James Heckman and coauthors in which they evaluate the economic outcomes of GED holders. On average, GED recipients have higher hourly wages than dropouts (and lower wages than graduates), as we would expect [11-12]. After controlling for measured differences in ability (via Armed Forces Qualification Test (AFQT) scores), Heckman and Rubinstein find that GED recipients earn less per year, have lower hourly wages, and complete fewer years of school than other high school dropouts [11]. If GED recipients' struggles in the labor market were due solely to lower cognitive abilities (perhaps partially due to fewer years of schooling), we would expect them to perform on par with (if not better than) other high school dropouts once AFQT scores were taken into account. The fact that their labor market outcomes are worse after controlling for differences in ability suggests that there are other unmeasurable factors that account for this difference; the authors surmise that the primary factor is less-developed noncognitive skills [11-12]. Specifically, GED recipients have more behavioral and personality problems than any other group; once these differences are controlled for, the wage differences are insignificant [12]. The fact that their behavioral differences are driving their wage differences suggests that noncognitive skills play an important role in determining labor market employability and advancement.

Noncognitive skills have been found to be important outside the labor market as well. Heckman and Kautz (2004), for example, find that character skills are a better predictor of educational attainment, labor market success, health, *and* criminality than are IQ scores [13]. Similarly, in an experimental study of the Perry Preschool Program, improved "personality skills" were found to significantly improve adult outcomes. Those in the treatment group, for example, had a lower likelihood of criminal activity and were less likely to suffer from long-term unemployment [14]. Those in the treatment group had 2.5-hour-long sessions, five days a week, in which "active participatory learning" was stressed. Children were taught (a) how to plan, execute, and evaluate tasks, (b) how to cooperate with others, and (c) the basics of interpersonal conflict resolution [14]. Results included an increase in academic motivation, greater cognition, and a reduction in aggressive, antisocial, and rule-

breaking behaviors (termed "externalizing behaviors") [14]. Long-term improvements in IQ, however, were not found [14]. Most important, noncognitive factors—including curiosity, self-control, persistence and conscientiousness—were found to be responsible for nearly two-thirds of the benefits children realized (specifically, these benefits included being more likely to graduate from high school, more likely to be employed at age 27, more likely to earn more than $25,000 per year by age 40, less likely to be arrested, and less likely to be on welfare than their peers) [15].

In another experiment, children were categorized into two groups based on their own perceptions of their math abilities [7]. Those who perceived themselves to be better at math (had higher math self-efficacy) performed better on average—that is, they reworked more problems, had more positive attitudes, and ultimately solved more problems [7]. These findings are likely tied to the fact that those with high self-efficacy have been found to be the most able to persist in the face of challenges. Finally, in a study of 140 eighth graders using questionnaires administered to students and their parents as well as standardized test scores, school attendance, hours spent doing homework, and selection into a competitive high school program, Duckworth and Seligman (2005) found that self-discipline is more important than IQ in predicting various measures of academic performance [16]. Specifically, highly self-disciplined students were found to outperform their peers on every academic performance variable [16].

The development of social and other noncognitive skills is a primary aspect of children's and adolescents' growth. As such, it features prominently in the growth experienced in school-aged children. That is,

> School is not only a place where children accumulate facts and learn academic skills, but it is also a place in which their basic motivation toward competence and achievement is established, their affiliative tendencies and relational patterns take root, their view of themselves as persons of worth and value develops, and their sense of the world as a safe or dangerous place is formulated. ([6], p. xix)

School is also the first place where students experience and grapple with others' perceptions of their abilities and overall persona. Teachers and peers therefore provide important inputs into students' development of their sense of self. As developed in 1902 by Charles Cooley, the concept of the "looking-glass self" is just this—that self-beliefs are primarily informed by observing how one is perceived by others [6]. As a result, parents, siblings, other family members, teachers, and peers all play a significant role in the definition of the "self." These self-perceptions of ability, of course, also inform confidence, which "become[s] instrumental to the goals [people] pursue and to the control they are able to exercise over their environments" ([17], p. 13). Thus, the noncognitive skills of self-concept and self-efficacy (where self-efficacy relates to a specific task, and self-concept is a more global measure [6]) become important in determining the risks people are willing to take, the challenges

they are willing to confront, and the resulting opportunities to which they make themselves available. As a result, self-fulfilling prophecies often emerge. Since "the self-efficacy beliefs students hold when they approach new tasks serve as a filter through which new information is processed" ([10], p. 754), they may be less likely to approach challenges in which they have low self-efficacy and as a result never gain the experience that would serve to increase their self-efficacy. Therefore, "those who enter adulthood…plagued by nagging doubts about their capabilities find many aspects of their adult lives aversive, full of hardships, and depressing" ([18], p. 184). Similarly, low senses of self-efficacy can cause youth to forgo pursuing career tracks that they view to be unattainable because they doubt their ability to succeed [18]. Having established the importance of noncognitive skill development for both academic and labor market outcomes, we now turn to a discussion of the factors that *determine* noncognitive skills, and whether age is among them.

# Noncognitive skill determinants: Is age among them?

There is certainly no shortage of studies evaluating the determinants of noncognitive skill formation and growth. Nor is there a consensus among them—largely because most studies focus on a specific age range, a specific noncognitive skill, and/or a specific contributing factor. As a result, many of the literature's findings are simply not comparable. A number of studies find that age is *not* among the primary determinants of noncognitive skill growth; another set suggests either that age itself is a primary factor *or* that the primary factor is something directly influenced by age (such as experience). In this subsection, we review both "classes" of literature. We begin by presenting those studies finding that non-age-related factors are most important and then turn to a discussion of those studies finding that age itself does, in fact, matter.

## Non-age-related determinants of noncognitive skills

We divided this class of literature into four general categories:

1. Those finding that family background and socioeconomic status are important in determining noncognitive skill development
2. Those finding that individual experiences are most important
3. Those finding that the situation in which one finds oneself is key
4. Those finding that other factors are of primary importance

We now discuss each of these groups in turn.

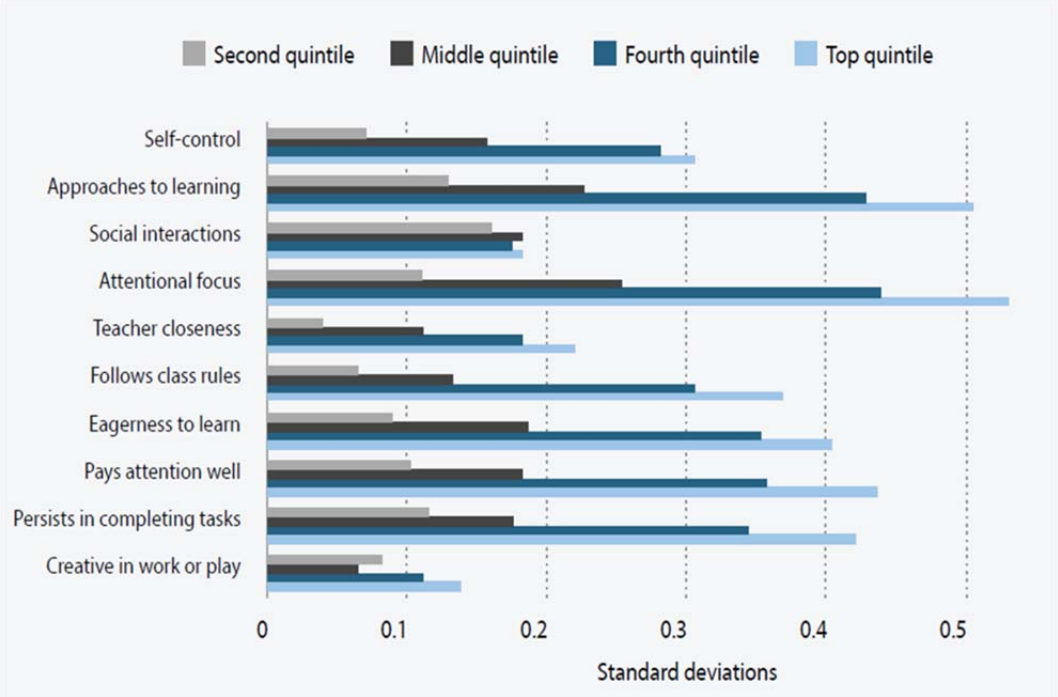
### *Family background and socioeconomic status (SES) matter*

A number of studies have pointed to the importance of SES and other family background characteristics (e.g., parental education, parental ability) in noncognitive skill development. Heckman, for example, finds that family environments (e.g., family income level, single parent or not, whether environment is supportive of children's academic growth) are important predictors of both cognitive and noncognitive abilities and that these environments also are important in predicting later-in-life outcomes, such as criminal activity and health [19]. He explains that "gaps in ability emerge early and persist"; most of the ability gaps present at age 18 existed at age 5 ([19], p. 298). And, most important, he provides a socioeconomically linked explanation for these early differences: "Middle class children receive massive doses of early enriched environments. Children from disadvantaged environments do not" ([19], p. 313). Thus, most of the ability gaps observed among young children can be tied to differences in the children's family's SES and their resulting "advantage" (or lack thereof) [19].

Cameron and Heckman (1998) find that family background and other age-nonvariant variables (e.g., parent's highest grade achieved and permanent family income) *maintain* their effects over time [20]. Specifically, a significant portion of the family-income/schooling relationship is due to permanent family/environmental factors, not to short-term liquidity constraints or other short-term factors [20]. This suggests that we are unlikely to find a key "age range" in which noncognitive skill development is most likely to occur because SES and other family background characteristics are not likely to drastically change at any particular ages. In a separate effort, Cunha and Heckman (2010) develop a theoretical model of skill development (both cognitive and noncognitive) based on findings from other studies [21]. In their two-staged model, investments into children's "skills" are made in period 1, period 2, or in both periods. Skills produced in period 1 increase the productivity of period 2 investments (a phenomenon called "dynamic complementarity") [21]. Their model showed that lower investments are made in children in credit-constrained families, not only in period 1, but also in period 2 [21]. As a result, it becomes inefficient to invest in these children in period 2 due to the lack of period 1 investments (and thus the lack of dynamic complementarity) [21]. Specifically, it would be more efficient to invest in the new generation's period 1 development. That said, a period 2 "ChalleNGe" investment might still be an effective way to improve the skills of those adolescents lacking in period 1 investments; it simply wouldn't be as efficient as those investments in those same children when they were younger. In this framework, the credit constraints and level of early investments ultimately determine people's adolescent and adult skill sets.

Other studies also find that SES is important in the generation of noncognitive skills. Garcia (2014), for example, finds a relative disadvantage among children in the lowest SES quintile in *all* major noncognitive domains analyzed [22]. This is illustrated in Figure 2. Although there is some variation across domains in the relative ranking of the first through fourth quintiles (highest to second to lowest), on average the quintiles are ranked in increasing order. For example, the second quintile has better self-control than the bottom quintile, the middle quintile outperforms the second, the fourth outperforms the middle, and the top outperforms the fourth. This suggests that SES is an important driver of noncognitive skill development.

Figure 2. Noncognitive skills gaps by SES quintile, as compared with the bottom quintile




Source: [22].

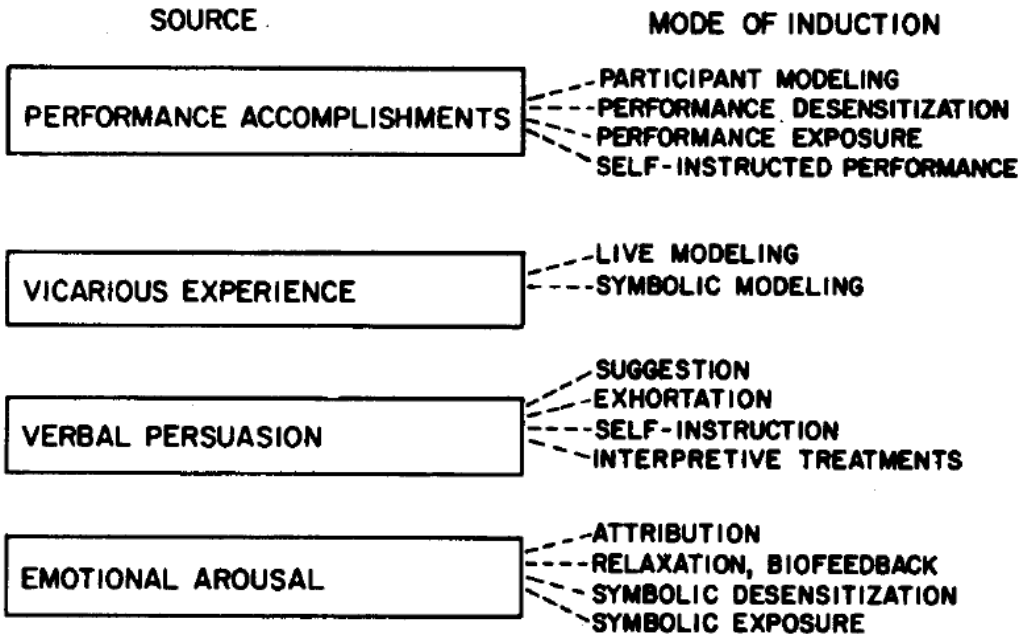
Similarly, Heckman's work on GED recipients illustrates that members of this group tend to come from disadvantaged backgrounds [12]. Since—as we discussed earlier—GED recipients also tend to be weaker in noncognitive skills than their high-school-graduate counterparts, this suggests that their disadvantaged backgrounds may have been at least partially responsible for their slowed development of noncognitive skills (as demonstrated by their failure to complete high school). Likewise, in a study evaluating the intergenerational transmission of both cognitive and noncognitive skills, Anger (2011) finds that, regardless of age, children with more educated

parents have a more internal locus of control [23]. And parental education, of course, is directly linked to SES. In addition to their education levels, parents' perceptions—specifically, their perceptions of their abilities to help their children succeed academically (and thus ultimately in the workforce)—have also been shown to directly affect their children's noncognitive development [18]. Parents are more likely to believe that they can help their children succeed if they are more educated, and thus come from a higher SES background.

### *Experiences matter*

In addition to family background, and specifically SES, noncognitive development is also influenced by personal experiences. Bandura (1977) posited that four major sources of information inform self-efficacy: performance accomplishments, vicarious experience, verbal persuasion, and physiological states [18]. Figure 3 presents Bandura's theorized inputs to each of the four information sources. Of the four sources, he finds that performance accomplishments are those that have the largest effects on efficacy beliefs [18].

Figure 3. Efficacy expectations




Source: [18].

Others have validated this finding. Coleman and DeLeire (2003), for example, find that stressful life events (experienced personally, not vicariously) contribute to a

more external locus of control and are more important than age [24]. Pajares (2005) also notes that mastery experience (another term for performance accomplishments) is the most influential source, especially when compared with vicarious experience, verbal persuasion, or emotional arousal [25]. He notes that experiencing success is important for building self-efficacy, whereas experiencing failure can result in a paralyzing sense of anxiety or dread due to a fear of additional failure [25]. In addition to experiencing previous success (or failure), a person's perceptions about the reasons for that success (or failure) also are important [26]. Specifically, "success enhances self-perceptions of competence only if the child accepts responsibility for that success" ([26], p. 130). Thus, locus of control is instrumental in determining how past experiences influence future actions or performance. If a child believes that the experience was contingent on his or her own behavior, this will make his or her locus of control more internal. If, however, the perception is that the success or experience was driven by external factors, the locus of control will likely become more external. These perceptions need not be directly correlated with age.

### *Situations matter*

In addition to previous experiences, many studies note that the situation in which one finds oneself at any given time will likely affect one's noncognitive skills. As explained by Rotter (1975), a person's expectations in a particular situation are determined both by previous experience in *that exact* situation and by more generalized experiences in situations that he or she might perceive to be similar [27]. Similarly, Cairns et al. (1990) evaluated the stability of locus of control and other aspects of self-concept in an effort to determine whether any of these noncognitive characteristics are situational in nature [28]. They do find evidence of situational variation. Specifically, over an 18-month period, they compare the stability of noncognitive measures for three populations: those who were in school full-time after age 16 and attended the same school, those who changed schools but still attended full time, and those who left school for full-time employment [28]. The authors find greater stability in self-concept over time for the two groups who remained in school, and those who stayed at the same school exhibited the most internal locus of control [28]. They also found that the overall changes due to age were small, perhaps suggesting that the "period of late adolescence is one of relative stability" ([28], p. 943). Thus, stability of situation seemed more important than age in determining noncognitive skill development.

We end this subsection with an example of how situation has been found to influence noncognitive skills and perceptions. Hendricks and Montgomery (1984) conducted a study of 48 unmarried adolescent fathers, together with 50 adolescents who did *not* father out of wedlock; the subjects were asked about their locus of control [29]. Those who had fathered out of wedlock in adolescence were more likely to feel that there was little they could do to change their lives; they felt that their probability of success was mostly a matter of luck [29]. Thus, they were more likely to have an

external locus of control than their peers. Unwed adolescent fatherhood, therefore, is one example of a situation that can affect an individual's sense of control.[2] The precise age at which fatherhood occurred did not appear to be important.

### *Other things that affect noncognitive skills*

A number of other studies find effects of non-age-variant characteristics on noncognitive skill development. Among these findings are the following:

- Noncognitive skills are malleable and affected by a number of factors, but notably not by age [30].
- Motivation (which may not vary with age) is an important determinant of achievement [31].
- Teaching approaches are important in developing noncognitive skills [32].
- Praise can have an impact on students' beliefs and motivations [6].
- Parental perceptions of their children's abilities affect how children perceive their own abilities [33].

## Age as a possible determinant of noncognitive skills

Having reviewed those studies finding that the primary determinants of noncognitive skill development are *not* age related, we now turn to a discussion of those that find that age *is,* in fact, important. Some of these studies point to age as having only an indirect role (e.g., age matters because acquired experience matters, or age matters because grade level matters). In this subsection, we begin with a discussion of the importance of age as a correlate with other, more important determinants (such as situation or experience, as we discussed in the previous subsection); we then review those studies finding that earlier interventions are important and that age independently matters for developmental processes.

### *Development throughout the school grades*

Children's and adolescents' development processes (both cognitive and noncognitive) have been found to correlate significantly with their progression through the school

[2] It may not be that the "situation" of adolescent, unwed fatherhood *caused* a more external locus of control to develop. It may, instead, be that the external locus of control was preexisting and caused the behaviors that led to the fatherhood situation. That is, these are correlations and do not necessarily imply causation.

grades. The middle school transition has been found to be particularly important and can detrimentally affect students' levels of noncognitive skills. The transition from the elementary to the middle school climate is one that reduces the closeness that students feel to their teachers (since they now have a different teacher for every subject) and students are graded in terms of how they perform relative to their peers (for many, this is the first time they experience being graded on a curve) [18, 34]. It is also the first time that students experience being assigned to a class or academic track based on their ability [34]. This is likely to increase not only students' competitive nature, but also the extent to which they look to others' abilities in making self-assessments.

In addition, in the early school grades, because students of all ability levels are in one classroom, the tasks that teachers assign are those they expect all students to be able to master [35]. This can result in an overestimated sense of self-efficacy among students [35]. On transitioning to middle school, they may lose the initiative that previously emerged from their anticipated ability to complete all tasks; they may also lose some "control" over their academic progress since the tasks are more difficult and students may not always succeed in completing them simply by working on them for longer [35]. The decreased sense of control over their academic progress that students experience as they leave the elementary school classroom, combined with the heightened anxiety over their performance and ability levels, may ultimately result in decreased academic motivation, especially for those students who do not perform as well as their peers [34]. In addition, the transition from elementary to junior high is accompanied by a change in educational approach: elementary school subjects are taught based on a "building blocks" approach, and students may be more confident in their ability to excel in a new subject or task since it very clearly builds on knowledge they have already mastered [34]. In junior high, however, for the first time, students are faced with "tracking," wherein students are grouped by ability. Thus, students may begin to doubt their perceived competence at new tasks, regardless of how well they have performed historically. This is yet another way in which students' sense of control over their academic outcomes and, hence, motivation may be affected by the school transition; as a result, they may experience a shift in their locus of control from more internal to more external.

After making the transition to high school, students begin to experience a heightened sense of control due to the increased social and personal freedoms that come with their age; thus, students then experience an increasingly internal locus of control [36]. In a longitudinal study of 174 high school students, Chubb and Fertman (1997) find that, in each year between 9$^{th}$ and 12$^{th}$ grades, adolescents' locus of control becomes more internal [36]. By the time students reach 12$^{th}$ grade and ultimately graduate, they are feeling their most empowered. These relationships between school grade and noncognitive growth suggest that there is also a role for age in determining noncognitive growth (and perhaps when it is most likely to occur) since

school grade and age are highly correlated. Thus, once again, age is important because situation (defined here by school grade) matters.

### *Development as a function of experience*

The literature also reveals that previously accumulated experience is an important determinant of noncognitive skill development. If we assume that, on average, the likelihood of having certain experiences increases with age, this is yet another *indirect* argument for the role of age. "Mastery experience"—a synonym for the "performance accomplishment" source of self-efficacy shown in Figure 3—has been found to be the most important source [17]. Students who have personally experienced academic and task achievement will likely have more confidence in their work. In addition, they will monitor their time more effectively, will be more efficient problem solvers, will be more persistent than their equally able peers, will work harder, and will evaluate their progress more frequently [10]. The likelihood of having sufficient mastery experience, one could argue, increases with exposure to varied environments and challenges, thus increasing with age.

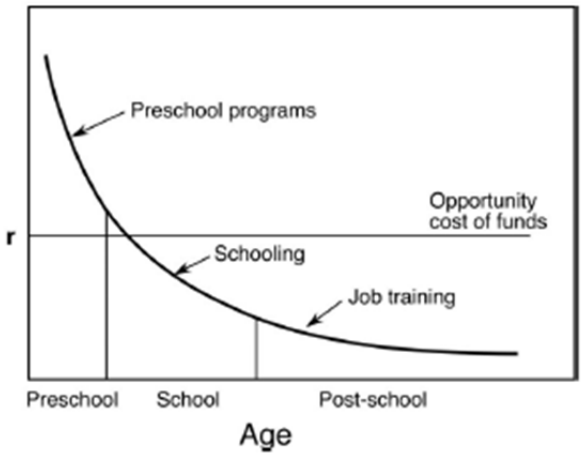
Previous experiences are especially important when students face unfamiliar tasks. In such cases, they are forced to draw on experiences that they view as similar [37]. As they age and acquire a greater variety of experiences, the likelihood increases that they will be able to think back on a situation in their history from which they can gain insights and confidence regarding how best to approach the unfamiliar tasks at hand. The greater students' previous experience in tackling either new tasks or in ultimately succeeding in a wide variety of situations, the greater their ability to anticipate a positive outcome in a *new* situation will be. This ability to project a positive expectation based on past experience is important because it provides students with an additional source of motivation [35].

### *Early investments are important*

A number of studies have also pointed to the importance of early investments in children's development—both cognitive and noncognitive. This is yet another way in which age has been found to be important for the effectiveness of noncognitive investments or interventions: the earlier they happen, the more effective they are likely to be. Numerous studies point to the dynamic complementarity of skill formation, in which later investments are most effective when able to build on earlier investments. Heckman (2000), for example, stresses the importance of early cognitive-skill investments since "more able people acquire more skills [and] more skilled people [then] become more able" ([38] , p. 6). These differences in investment returns by age are also found when considering the noncognitive aspects of development [39]; Figure 4 illustrates the decreasing returns (and changes in the typical types of intervention) as children age.

Similarly, Cunha et al. (2006) view childhood as a multistage process, in which early investments serve as an input into the productivity of later investments [40]. They therefore suggest that, if early investments are going to have a sufficient return, they *must* be followed by later investments [40]. Smith (1999) evaluates social capital formation in adolescents and whether it affects their civic participation in young adulthood [41]. She also finds evidence of dynamic complementarity in that parental involvement is necessary to help build students' 8th grade social capital but is less important in building the stock of 10th grade social capital, as this is grown from the "base" created in 8th grade [41]. Within this framework, all policy should focus on maximizing early-childhood investments, under the assumption that later investments will never reach their full potential without these early investments. That said, regardless of the age at which investments are made, there is always a positive return; it is simply smaller at later ages. Thus, although restricting the ChalleNGe population to 16-year-olds would be the best way to maximize the return on investment, there would still be positive gains to be made in the 17- and 18-year-old populations.

Figure 4. Rates of return to human capital investments




Source: [42].

While economically efficient, the authors' suggested approach might ultimately result in neglecting adolescents and young adults who did not receive what the authors would consider the "necessary" early investments. Not only is this likely not a realistic policy approach, it also would not work in the ChalleNGe framework. It would suggest that ChalleNGe should focus solely on the 16-year-old population (the youngest population they can influence), so that the investments ChalleNGe makes in these cadets will have the greatest maximum potential and can be amplified by later investments. This is unrealistic for a number of reasons. First, the 17- and 18-year-

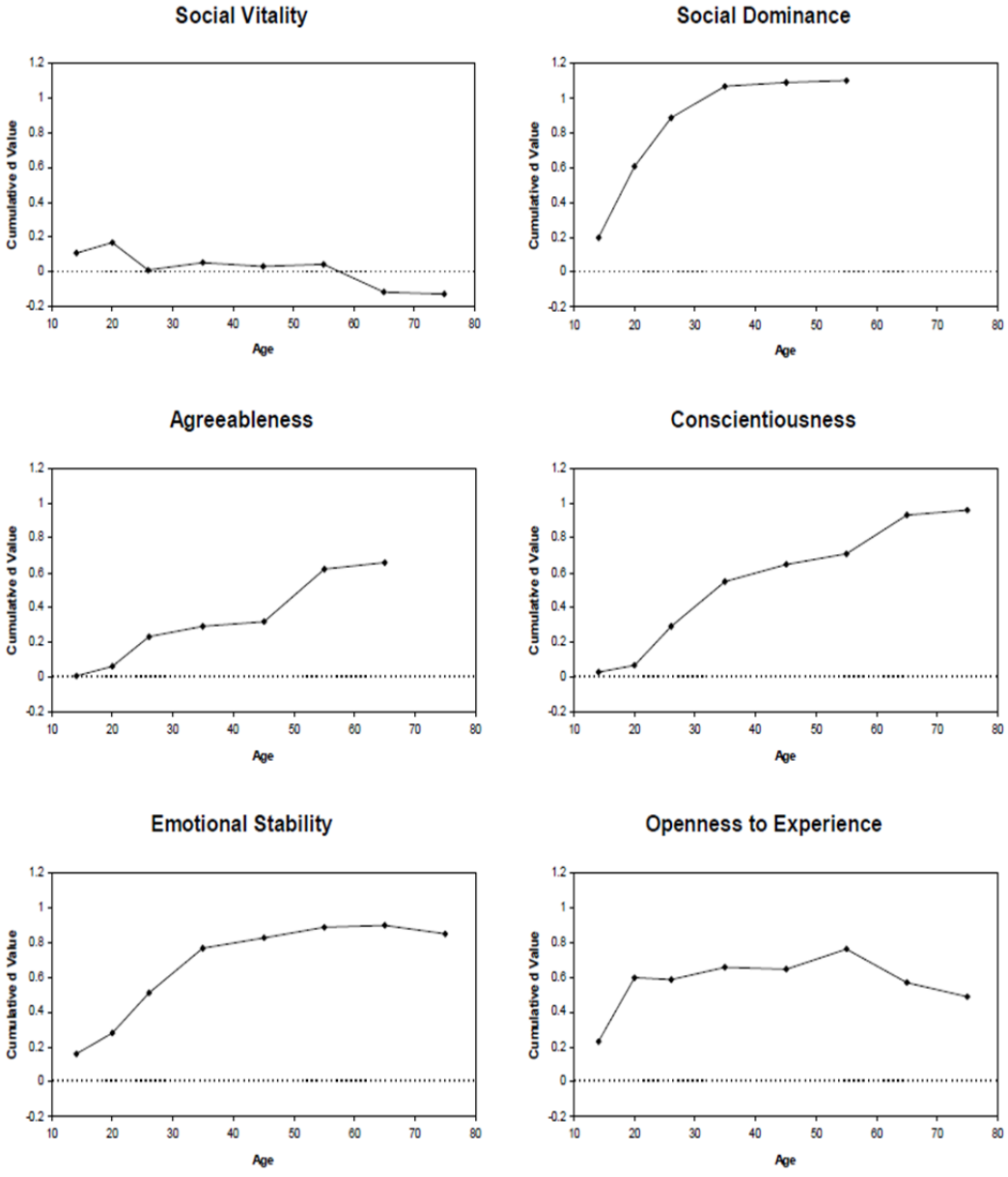
olds are an important element of the ChalleNGe cohorts; with greater experience, they can serve as leaders and mentors for their 16-year-old peers. Second, older cadets are more likely to have earned enough high school credits before their admission to ChalleNGe to make credit recovery or the earning of a high school diploma while at ChalleNGe realistic goals; thus, if limited to 16-year-olds, ChalleNGe would become a largely GED-granting program. Finally, limiting admissions to 16-year-olds also would mean that ChalleNGe would be ignoring a large portion of the at-risk population solely on the basis of their age, a decision the ChalleNGe leadership is likely unwilling to make. The important takeaway for ChalleNGe is that investments are most productive the earlier they are made; the program should consider all of the implications of excluding 16-year-olds because it might have the longest term impacts on this population.

### *Age as an independent determinant of noncognitive development*

Having reviewed those studies concluding that age plays an indirect role in noncognitive growth, we now turn to a discussion of studies finding that age has a more *direct* role. It is worth noting that most studies on noncognitive growth simply do not focus on age and, among those that do, even fewer include age ranges that are relevant to ChalleNGe (e.g., a study of sense of control among 18- to 50-year olds and those age 51 and older [43]). We focus in this subsection on those studies that concern populations relevant to ChalleNGe: we begin by reviewing those that include the ChalleNGe age range among their ages of focus. We follow this by a discussion of studies focused on adolescents.

A few studies have attempted to establish an understanding of how noncognitive skills evolve over the life cycle. Roberts et al. (2006), for example, use the Big Five taxonomy (extraversion, agreeableness, conscientiousness, emotional stability, and openness to experience) to synthesize and compare results from a number of other studies [44]. They find that many of the personality changes that are related to age result from the timing of biological changes as well as changes in "normative life tasks and roles, such as leaving home, establishing a family, and starting a career"—all changes that occur for most in young adulthood but may happen earlier for the at-risk ChalleNGe population ([44], p. 2). The personality traits for which they find significant increases in adolescence are social dominance and emotional stability [44]. For these and the other measured traits, as shown in Figure 5, they find that the most drastic changes occur during young adulthood (20 to 40 years of age), not during adolescence [44]. They also, however, argue that psychological age may be more important than chronological age, and that the age at which people leave their parents' homes and begin to form their careers and/or families is when the most significant noncognitive growth will be observed [44]. Although many noncognitive changes have been thought to be tied to age, this suggests that they may, in fact, be more related to situations and experiences.

Figure 5. Cumulative mean-level changes in personality across the life cycle[a]




Source:[44].

[a.] The cumulative d values on the y-axes represent the total lifetime change in standard deviations.

The primary reason for the literature's expectations that most changes would occur in adolescence is the complexity of that period of life. As explained by Wigfield et al., "very few developmental periods are characterized by so many changes at so many different levels as adolescence" ([45] p. 32); adolescents simultaneously experience pubertal development, social role redefinitions, cognitive development, school transitions, emerging sexuality, and changing familial relationships [45].

Not only must adolescents grapple with a large number of such simultaneous changes, but there is also evidence that their thinking changes. Specifically, they have been found more able to think abstractly, to consider hypothetical (as well as real) situations, to more elaborately process information, to consider multiple dimensions of a problem at once, and to engage in self-reflection [6]. As such, they are likely developing a sense of self-concept for the first time and arguments could be made that, the younger in adolescence any interventions occur, the more likely they are to be effective, prior to other sources establishing a firm influence. Adolescence is also important because the maturation of the prefrontal cortex develops during adolescence—later than all other parts of the brain [46]. The prefrontal cortex is the area of the brain that controls executive functioning—the mental processes that enable us to plan, focus attention, remember instructions, and juggle multiple tasks successfully. It also is linked to social cognition, or a person's ability to see the perspectives of others [46]. This is especially important for adolescents' noncognitive development, as they develop for the first time a sense of how they are viewed by others (and similarly how they compare with others).

In addition, Schunk and Pajares (2002) find that the 12- to 16-year-old age range is a key period of influence, and also a period when parental involvement tends to decline [35]. Thus, this is an ideal time for external sources to influence noncognitive development, suggesting that ChalleNGe should keep its hold in the 16-year-old market. And although earlier studies we discussed have found that it is essential that cognitive investments be made early in children's lives, since later cognitive investments do not serve as good substitutes for early investments, this is not necessarily true when considering noncognitive skills [47]. Cunha et al. (2010) find that the complementarity between noncognitive skills and prior investments become weaker as children age, making it easier at later ages to "remediate early disadvantage using investments in noncognitive skills" ([47], p. 921). Thus, noncognitive deficits can be overcome at later ages than cognitive skill deficits, making age somewhat less important in determining an optimal time for noncognitive growth intervention than it is for cognitive growth.

# Recommendations for ChalleNGe

Here, we extrapolate recommendations for ChalleNGe—and, specifically, whether ChalleNGe should consider further restricting the qualifying age range—from the

literature we have reviewed. Among the most important findings for ChalleNGe to consider is the overall greater malleability of noncognitive than cognitive skills (and thus the less important it is that noncognitive growth be achieved by a certain age since noncognitive skills can be effectively influenced at later ages, unlike cognitive skills). In addition, those age ranges in which most noncognitive growth is found to occur are outside the 16- to 18-year-old range of the ChalleNGe cadets. Thus, further limiting cadets' ages will likely not have a significant impact on noncognitive skill growth.

Much of the literature highlights the fact that noncognitive skills are more malleable throughout life than are cognitive skills. Specifically, although "pure IQ is stubbornly resistant to improvement after about age eight…executive functions and the ability to handle stress and manage strong emotions can be improved, sometimes dramatically, well into adolescence and even adulthood" ([48], p. 48; [38]). Noncognitive skills change over the life cycle largely because of changing experiences, situations, biology, and investments [49], and noncognitive skill development is less dependent on previous investments than is true of cognitive skill development. The fact that noncognitive skills are malleable and thus open to influence from new experiences and perspectives suggests that the precise age at which cadets arrive at ChalleNGe will not be that important in determining the feasible growth in these skills.

Although some studies do point to specific age ranges as the most productive or most likely time for noncognitive development, these ages aren't particularly relevant to ChalleNGe. Thus, where ChalleNGe can be most helpful is in aiding noncognitive growth for those who arrive at ChalleNGe with a noncognitive skill deficit relative to their peers. The majority of these studies find that noncognitive skill development primarily happens either in early adolescence (ages 13 to 15) or in early adulthood (ages 18 to 25) [47]. There is not, however, sufficient evidence in differences in noncognitive development among 16-, 17-, and 18-year-olds to warrant a change in ChalleNGe's age policy. A number of studies *do* indicate that the earlier noncognitive interventions are started, the more likely they are to be effective. This would suggest that accepting more 16-year-olds would be optimal. There are a number of other factors, however, that ChalleNGe should consider before changing the age-mix of its cadets. These include the fact that (a) decreasing the population of 17- and 18-year-olds would decrease the population of cadets who can serve as mentors or leaders to their younger counterparts and (b) 16-year-olds may find it more difficult to make their transition to the civilian economy after completing ChalleNGe because they are not legal adults and may have a harder time finding employment. Overall, since noncognitive skill development is dependent on a number of independent factors, we do not recommend further restricting the qualifying age range in an effort to maximize cadets' noncognitive skill growth.

# Initial TABE Scores and Potential for Score Growth: Is There a Relationship?

Having completed our review of studies addressing the determinants of noncognitive growth and whether ChalleNGe should consider changing its age qualifications based on when noncognitive growth is most likely, we now turn to a similar policy question regarding implementing a minimum cognitive (TABE score) requirement. The specific question we were asked to investigate was whether cadets with certain TABE scores are most likely to experience significant TABE improvement. And, if so, what are those scores?

The following limitations in the existing literature required us to alter this question somewhat to broaden its scope:

1. There is little to no literature on the TABE. Existing research either addresses standardized tests as a whole or focuses on more familiar exams, such as the Scholastic Assessment Test (SAT), American College Test (ACT), or Graduate Record Examination (GRE).

2. Within the standardized testing literature, there are no studies that identify the potential for test-score improvement for a given initial score—likely because of the wide number of factors that influence score improvement.

As a result, we focused on the potential trade-offs and unexpected consequences that ChalleNGe should consider when determining whether to introduce a minimum TABE score for admission (and thus placing greater emphasis on the role of cognitive scores). As part of this effort, we also reviewed studies that offer recommendations on the most effective ways to prepare students for standardized tests, to administer them, and to evaluate results, in case ChalleNGe *does* implement a minimum score. Adopting a minimum score requirement would increase the program's reliance on and the importance of test scores, so we also address the following three questions:

- What has the literature found regarding the ability of test scores to predict success?

- What else might test scores be capturing (in addition to academic ability)?

- What are potential unexpected consequences for ChalleNGe of relying more on test scores as a metric of cadets' cognitive improvements?

# Standardized exams as predictors of success

We begin with a brief discussion of the existing literature's findings on the ability of entrance exams to accurately predict future success. This question is important because it reflects the ChalleNGe program's current dilemma; if it uses the initial TABE scores to determine whether a particular high school dropout is admitted into the program, it is essentially using the TABE as an entrance exam. If entrance exams are not good predictors of how well a student does academically (or in other measures of success), this suggests that the ChalleNGe program should carefully consider all repercussions before implementing a TABE-score admission cutoff.

In a study evaluating whether community-college placement tests—which are used to determine the appropriate course level for incoming students—are highly predictive of subsequent course grades, Armstrong (2000) found that the relationship was weak at best and that a number of high school performance variables were better predictors [50]. He notes that "sorting students by using cutoff scores on a test may mask important individual student characteristics and situations" ([50], p. 682). Specifically, he finds that student indicators, not placement test scores, are most useful in predicting course outcomes. These indicators include their high school grade-point average, the grade in their last English or math class, and the number of years of English and math that were taken in high school [50].

In a related study, Scott-Clayton (2012) evaluated the accuracy of community-college placement exams in predicting students' course grades [51]. She used three measures of "success": whether the student earned a B or better in the first college-level course in the relevant subject, whether he/she earned a C or better, and whether he/she passed (earning a D- or better). She found that placement tests predict who is likely to do well in a course more accurately than they predict who is likely to fail a course [51]. This suggests that placement exams might not be that relevant in the ChalleNGe construct since the program is most interested in implementing minimum test-score requirements as an indication of who is not likely to complete the program (i.e., who is most likely to fail).[3] Most important, Scott-Clayton suggests that using placement exams as a screening mechanism can be more problematic than not screening at all: the increase in the number of otherwise-qualified students prevented from accessing

[3] In our analysis of the cadet survey data (in an upcoming report), we will analyze the effects of implementing a minimum admission score. Specifically, we will determine implications in terms of "un-served," at-risk youth if different test-score criteria are imposed.

college-level courses because of a low test score exceeded the number of unqualified students admitted into college-level courses (who subsequently failed) [51].

## Possible test score pollutants

In the previous subsection, we presented an argument against implementing a minimum test score requirement at ChalleNGe, based on the inability of placement exams to predict course success. In this subsection, we turn to a discussion of *why* it is that test scores may not accurately capture students' abilities. The primary mechanism through which this occurs is what the literature refers to as "test-score pollution." Put simply, there are factors *other* than cognition or ability affecting test scores. As a result, the information one hopes to extract from test scores (about math abilities, for example) becomes polluted. A few of the common culprits of test-score pollution include proficiency in the English language, economic environment, family mobility, anxiety, stress, fatigue, and motivation [52]. Economic environment and family mobility, for example, and overall socioeconomic status, may affect test preparation and confidence. Noncognitive skills can be important for test taking as well. In particular, a well-developed sense of efficacy in a subject results in students feeling more confident in their abilities and thus being less likely to second-guess themselves while taking a test [53].

Anxiety and stress tend to be most prevalent in testing situations that are "high stakes"—that is, when the students understand that their test results will be important in determining future outcomes. For this reason, the SAT is a particularly anxiety-producing test since it is viewed as a critical step in college admissions [54]. Similarly, if ChalleNGe applicants took the TABE knowing they had to score high enough to be admitted into the program, anxiety levels could feasibly increase, thus decreasing the validity of the test scores. Evidence of anxiety-polluted test scores has been found in the prison environment: "These [anxiety] states are maximal at prison intake, when the offender is struggling to adjust to dramatic changes in daily life" ([55], p. 239). Although prison is clearly different from the ChalleNGe environment, there are similarities to be found in the sudden and drastic lifestyle changes. Piccone (2006) finds that the TABE score improvements of 18.5 percent found among inmates is simply infeasible, suggesting that anxiety is overriding the prisoners' abilities to fully apply themselves to the TABE at intake [55]. If ChalleNGe cadets experience similar levels of anxiety in the first few days (or even weeks) at ChalleNGe, the validity of initial TABE scores could be brought into question. In addition, we have anecdotally heard from ChalleNGe directors that cadets may not take the initial TABE seriously; many of the cadets, at that stage, are not very motivated, view tests as part of the educational system that has already failed them, or simply do not want to be at ChalleNGe. Thus, the ability to make decisions as to

which cadets are likely to succeed at ChalleNGe (and thereafter) based on a pre-admission TABE placement exam would be questionable at best.

In the remainder of this subsection, we discuss three common test-score pollutants—the things that get muddled with ability to produce a test score that may not be accurate. It is impossible to separate a student's ability from these factors: (1) classroom and teacher effects (both of current classrooms/teachers and previous ones), (2) stereotype threat, and (3) teaching to the test.

## Previous classroom and teacher effects

Some studies have highlighted the importance of the classroom environment in determining students' confidence, motivation, and ultimate success. Classroom environment—to include teachers' quality and effectiveness—might then directly affect test scores. This is especially true if the students have learned how to reason their way through difficult problems and are able to maintain confidence when faced with difficult academic tasks. An encouraging classroom environment is not only important for students' *current* development, but it also can have an impact on students' approach to learning in the distant future. Using data both on test scores and on classroom/teacher assignment in grades 3 through 8, Chetty et al. (2011) found that teacher quality is important, not only in improving test scores, but also in determining longer term outcomes; students with higher quality teachers were less likely to have a child as a teenager, lived in higher quality neighborhoods as adults, and had higher 401K retirement savings rates [56]. The authors also found other evidence of long-term impacts of teacher quality: test-score gains were still maintained three years after having a high-quality teacher, and students of higher quality teachers were more likely to attend high-quality colleges [56]. Similarly, Goldhaber et al. (1999) found that unobservable school, teacher, and classroom characteristics are significant predictors of 10$^{th}$ grade math achievement, as measured by test scores [57]. Their effects are not inconsequential; these three classes of characteristics, combined, account for 25 percent of the total variation in 10$^{th}$ grade math achievement [57]. Thus, the particular teachers, classrooms, and schools to which a student is assigned can have significant developmental and noncognitive impacts; these are precisely the types of effects that pollute test scores, decreasing their interpretive value.

## Stereotype threat

Stereotype threat is another well-documented test-score pollutant. Among a population of stereotyped or stigmatized individuals, stereotype threat is the sense that one might inadvertently take actions that would confirm an existing stereotype [58].

Broadly speaking,

> Stereotype threat occurs when individuals identify with a group (positive association between self and group) and identify with an ability domain (positive association between self and ability domain), while a negative stereotype suggests a negative connection between one's group and the domain at hand ([59], p. 3).

For example, the "group" with which an individual identifies might be "women" and the ability domain could be "strong math skills." The existing negative stereotype, however, that women have lower quantitative abilities, would suggest that the group and domain should, in fact, be negatively related (women should be less likely to perform well in math). In such cases, individuals may become "threatened" by the existing stereotype; if they act in ways that conform to the stereotype (e.g., by performing poorly on the TABE), they make "the stereotype more plausible as a self-characterization in the eyes of others, and perhaps even in [their] own eyes" ([60], p. 797). Any frustration or confusion felt during a test might alert the test-taker to this possibility—that if they do not perform well they will be confirming a negative group stereotype [58]. As a result, students become distracted and are less able to focus completely on the test material; they are therefore likely to perform worse than their true cognitive abilities would dictate. This is how stereotype threat serves as a test pollutant. Test scores are biased downward by the presence of stereotype threat and no longer serve as a true reflection of ability.

There are a number of ways in which stereotype threat can ultimately result in lower test scores. First, as mentioned above, "individuals may suffer negative performance outcomes" ([61], p. 647). The stereotype threat can also manifest itself as a form of anxiety and has been shown to result in a physical response that decreases test performance (due to the distraction) [59-60]. In addition, the stereotypes, coupled with academic struggles, raise the possibility that these struggles might be the result of internal flaws (i.e., that the individual simply is not capable of succeeding on that task [61]). A vicious cycle results that both causes and perpetuates the poor performance: "self-blame for a negative outcome leads to increased anxiety, in turn resulting in poor performance and even more self-blame" ([61], p. 649).

Over time, the "pejorative interpretations of failure" that the stereotype facilitates lead to increased stress and self-doubt, "diminish[ing] their sense of belonging to the academic arena" ([61], p. 647). As result of such experiences, stereotype-threatened students may become incentivized to redefine their self-concept so that the importance of academic achievement is minimized and is no longer part of one's personal identity [60]. This serves as a self-defense mechanism—since failure to succeed academically will no longer be damaging to one's self concept and stereotypes will no longer be "threatening"—but it has an unfortunate byproduct of decreased interest in school, decreased motivation, and further decreases in achievement [58, 62].

The importance of stereotype threat as a potential inhibitor to academic success has been illustrated in a few experimental studies. In the most well-known experiment of stereotype threat, Claude and Aronson (1995) administered a 30-minute section of the verbal Graduate Record Examination to black and white college students [60]. The verbal questions on the GRE are difficult enough to be at the limit of most college students' skills [60]. Those students in the "stereotype threat condition" were told that the test was diagnostic of their intellectual ability; those in the other condition were told that it was a problem-solving task that was not diagnostic of ability and were encouraged to view it as a challenge [60]. For some students, stereotypes were activated before taking the GRE section. They were asked to perform a word-fragment completion task in which words were missing letters [60], and 12 of the 18 word fragments had as one possible solution a word reflecting a race-related construct or image associated with African Americans [60]. The authors suggested that it was not necessary that the stereotype be believed; in the face of frustration, it may become more plausible as a self-characterization and thus more threatening [60]. Overall, blacks in the diagnostic condition performed worse than the nondiagnostic blacks, the diagnostic whites, and the nondiagnostic whites [60]. Black participants in the diagnostic condition also produced more race-related and self-doubt-related words than any other group [60]. The authors conclude that black underperformance is not, in fact, rooted in the demographic group or societal conditions. Rather, it is "a social psychological predicament of race, rife in the standardized testing situation" ([60], p. 810). This suggests that standardized testing environments are likely to evoke stereotype threat, which will negatively affect performance and make test scores less informative of overall ability than they are intended to be.

In addition to African Americans' verbal skills, the other subject area frequently discussed regarding the impact of stereotype threat is women's math abilities. In a more recent experiment, Steele, Spencer, and Aronson (2002) administered a 25-minute section of the math GRE [58]. Some women were given a "stereotype-threat reduction" explanation before taking the test; they were informed that, on average, men and women performed equally well on this test [58]. These women performed as well as their male counterparts, whereas those not given this explanation lagged their male counterparts [58]. The authors suggest that women taking a math test (and African Americans taking a literature or verbal ability test) might be trying to suppress the stereotype in the midst of the test, which detracts from test performance [58].

Stereotype threat is an important factor for ChalleNGe to consider in deciding whether to implement a minimum-qualifying TABE score. Not only is it most prominent in testing situations [59], but those tests with "distractingly important consequences" are the ones most likely to cause stereotype threat to emerge ([58], p. 2). This was illustrated in an experiment in the early 1960s, in which African Americans performed better on an IQ subtest when it was presented as a test of

hand-eye coordination—"a nonevaluative and thus threat-negative test" ([60], p. 798)—than when they were told it was an intelligence test [63]. If the TABE were administered to determine ChalleNGe qualification, it would become a test with important consequences. Test-takers would likely be distracted by the pressure to perform well and would contend with stereotype threat, decreasing the likelihood that the TABE results would accurately reflect their cognitive abilities. Stereotype threat has been noted to apply to any "academically stigmatized individuals" ([61], p. 647). ChalleNGe cadets (or potential cadets), as high school dropouts, most certainly fit this category. Their academic abilities and capacity to stick to a task and complete it are likely doubted by others; any academic failure would confirm the existing stereotypes.

## Teaching to the test

The final test-score pollutant we discuss is "teaching to the test," generally defined as altering class content so that it better aligns with the material on exams, thereby raising average test scores. Our sponsor's question—whether cadets with certain TABE scores are most likely to experience significant TABE improvement and, if so, what those scores are—suggests that the program is interested in maximizing TABE gains. An increased emphasis from ChalleNGe leadership on increasing cadets' TABE scores may result in consequential cultural changes among the teachers and staff, in an effort to maximize cadets' TABE gains, thus increasing their *final* TABE scores.[4] Although this test-score pollutant will not affect initial test scores—the mere act of introducing an admissions test will not result in teaching to the test—it likely will affect final TABE scores and thus TABE gains.

If increased importance is placed on students' test scores (and test-score improvements), teachers may become incentivized to alter their teaching approaches and curricula to maximize students' test score gains. As explained by Haladyna et al. (1991), the resulting test-score pollutants "increase or decrease test performance *without connection to the construct* represented by the test, producing construct-irrelevant test score variance" ([52], p.4). That is, students' test scores will no longer necessarily be a reflection of their inherent ability; they could also capture teachers' efforts to improve test scores (for example, via an increase in the number of practice tests or an adjustment of class content to more closely align with tested material).

The general notion of preparing students for tests—whether a test of course content or a standardized test intended to capture aptitude—is not, necessarily, always

[4] For example, at present, the program is heavily focused on improving cadets' noncognitive skills; a focused effort to increase TABE scores might come at the expense of noncognitive skill gains.

inadvisable. Both the content of instruction and the content of tests, for example, should be related to educational objectives [64]. Thus, instruction that is focused on achieving educational objectives will likely overlap with test content, as would be expected. There is a fine line between legitimate and illegitimate test preparation, and it is unclear where this occurs. Mehrens and Kaminski (1989) suggest the following continuum of legitimate to illegitimate instruction [64]:

1. General instruction on objectives not determined by looking at tests
2. Teaching test-taking skills
3. Instruction on objectives generated by a commercial organization that may have looked at the test
4. Instruction based on objectives that match those on the test
5. Instructions on specifically matched objectives
6. Practice on parallel form of same test
7. Practice on the same test

The authors note that the "crossover" from legitimate to illegitimate test-preparation techniques likely occurs somewhere between items 3 and 5, but argue that a precise determination cannot be made because such a determination will be situation specific [64]. Shephard (1990) suggests that the important distinction is whether "repeated practice or instruction [is] geared to the format of the test rather than the content domain" ([65], p. 19).

When teaching to the test occurs, there are significant consequences for students. Teachers are incentivized to focus their instruction on those areas that will be tested; this happens at the expense of nontested subjects and material (to include noncognitive skills) [65]. In addition, test scores eventually become inflated—potentially representing the teachers' classroom emphasis on test-taking "tricks" and tips rather than a greater understanding of the content being tested [65]. "Score increases may represent real gains in performance on tested parts of the domain, but they are misleading as estimates of improvement in the domain as a whole" ([66], p. 768). Students with inflated test scores ultimately suffer; test scores meant to reflect areas in which they need remedial help or extra attention no longer reveal this information.

# Recommendations for ChalleNGe

Based on the literature we reviewed, we do not recommend that ChalleNGe implement a minimum TABE score requirement for ChalleNGe participation (in order to maximize the TABE improvement of its cadets). Due to the number of test-score pollutants that cannot be separated from actual ability—such as anxiety, English-language proficiency, motivation, and family background—the overall validity of test scores and the extent to which they capture inherent ability is called into question. As a result, students with sufficient aptitude to thrive in the ChalleNGe environment and experience significant cognitive gains may, nonetheless, have standardized test scores that are much lower than their true abilities. If a test-score cutoff were put into place, this population would be denied the ChalleNGe experience. In addition, implementing a TABE score admission requirement would change the purpose of the program—from one designed to assist at-risk youth to one designed to help those at-risk kids who also score above the cutoff. A significant portion of the at-risk population would likely be removed from the population ChalleNGe is able to help.

It also may be more effective for ChalleNGe to maintain its focus on both noncognitive and cognitive skill improvement, in lieu of transitioning to a construct with greater emphasis on cognitive abilities. This is both because of the questionable ability of test scores to reflect true cognitive abilities (due to possible pollutants) *and* because noncognitive investments are more fruitful for adolescents. As children age, the types of investment with the largest returns shift—from cognitive investments early on to noncognitive investments in later years [67]. Given that ChalleNGe focuses on an adolescent/early adult population, investments made in developing and improving noncognitive skills will be more effective than investments that focus on cognitive skills. In addition, to meet the economic, political, and social demands of the labor force, cadets will need more than cognitive proficiencies. They will need interpersonal skills to be able to relate to others and will need intrapersonal skills to effectively evaluate and meet their own needs) [9].

If ChalleNGe leadership decides to implement a TABE-score admission cutoff and thus place greater emphasis on maximizing test-score gains, it will be important to carefully determine how to do so most effectively. The programs will need to ensure, for example, that important course content, such as noncognitive skill development, does not get removed in the aim of focusing only on TABE material. It will be important for teachers to carefully navigate the "fine line" discussed earlier between test preparation that is in line with educational objectives and preparation that is focused on teaching test-taking tricks. It also will be important that cadets understand the purpose and importance of the tests they take at ChalleNGe. They should be informed, for example, that having an accurate ASVAB score can be beneficial even if they do not have plans of enlisting in the services. Otherwise, it will

not be clear whether test scores represent cadets' true abilities or a lack of motivation to apply themselves to the tests.

It also will be important for ChalleNGe to minimize the possible role of other test-score pollutants. At a minimum, the program needs to recognize that students' experiences in previous classrooms and with previous teachers will affect their aptitude at any given time. Thus, students who perform poorly on aptitude tests may have been previously disadvantaged by having less effective or less engaged teachers; the score will not necessarily reflect "potential" aptitude as much as "current" aptitude. In addition, teachers of higher ability cadets will likely not see as much cognitive improvement in their classrooms as those with lower ability cadets (since there is simply less room for growth) [68]. Thus, if the programs become focused on maximizing cognitive gains and eventually go so far as to incorporate this in their teachers' evaluations, it will be important to keep in mind that the achievable growth will depend on cadets' aptitude when they arrived.

Finally, ChalleNGe will have to effectively combat stereotype threat if it is to increase the importance of TABE scores within the program. A number of possible techniques have been proposed in the literature. These include changing students' mindsets. First, it is important to get students to shift any blame for failing at academic tasks from a lack of intelligence to the difficulty of the particular context (in other words, moving them from a pejorative to a nonpejorative reaction to their academic challenges) [69]. Another outlook alteration important to overcoming stereotype threat is that students come to adapt an "incremental theory mindset" [70]. To quote Aronson et al.,

> [while] entity theorists are more likely to blame their own shortcomings for academic difficulties of failures…[making] internal, stable attributes for negative outcomes…incremental theorists interpret a struggle or a failure as indication that they didn't apply enough effort or approached a problem from the wrong angle. ([70], p. 658)

Similarly, presenting tests as being about evaluating "learnable skills" as opposed to the "limitations of ability" should reduce the amount of stereotype threat experienced ([58], p. 394). Finally, getting students to view intelligence as being malleable, as opposed to fixed, decreases the sense of limitation they will feel from their previous achievements or characteristics [58].

Although we do not recommend that ChalleNGe use TABE scores (or any test scores) as a cutoff for admissions—for the reasons discussed in this section—we view trying to improve cadets' ability to perform well on standardized tests as a perfectly acceptable objective. This is especially the case since standardized tests are the "preferred standard for college admission" ([50], p. 646). Because this is the case, any improvements in cadets' abilities to test well (on standardized tests, in general) could

help those returning to high school with potential college admissions in the future. They would also potentially help improve cadets' performance on the GED and better prepare those returning to high school for the testing element of achieving a diploma.

# Conclusions

The ChalleNGe program has recently been considering two changes to its admission criteria, in an effort to maximize cadets' skill growth, both cognitive and noncognitive, during the program's residential phase:

1. Further restricting the qualifying age range, to maximize noncognitive skill growth

2. Implementing a minimum required TABE score for admission, to maximize potential cognitive gains

Our future work will evaluate these questions analytically—using data collected at our seven-site cadet survey. For this effort, we have reviewed the existing literature on these two topics. Our synthesized findings suggest that neither of these policy changes is recommendable.

First, most noncognitive growth occurs outside the 16-18 age range (the current range of ChalleNGe cadets). Early investments are important, and earlier adolescent interventions have been shown to be more effective in spurring noncognitive development. In addition, no studies find a *direct*, independent impact of age on noncognitive growth. That said, there are noncognitive changes that are correlated with school progression (advancing through the school grades) and others associated with "mastery experience." For example, adolescents who have personally experienced challenges and been able to overcome them will have stronger noncognitive skills (including determination, self-concept, and locus of control). Thus, although there is no direct link between age and noncognitive development, noncognitive skills will vary with age because of the direct roles played by school progression and mastery experience. Based on these findings, restricting ChalleNGe to 16-year-olds would be most likely to maximize noncognitive growth since the 16-year-old cadets are unlikely to have acquired sufficient experience to have well-developed noncognitive skills. However, 16-year-olds leaving ChalleNGe often have the most difficulty being "placed": many employers will not consider them, and they are too young to enlist. In addition, there is a significantly sized 17- and 18-year-old at-risk population. If ChalleNGe were to focus its efforts on only 16-year-olds, it would be at the expense of these older at-risk adolescents.

A similar argument can be made against introducing a minimum test-score requirement. That is, requiring a minimum initial TABE score would remove all those

scoring below that cutoff from the possibility of receiving ChalleNGe's services. There are other reasons why a test-score admission requirement is not advisable. First, test scores have been found to be significantly less predictive of success (whether in terms of schooling, employment, or other outcomes) than other measures that also incorporate noncognitive skills. Thus, it is unlikely that initial TABE scores would be sufficiently predictive of achievable TABE gains or program completion. In addition, test scores are riddled with a variety of pollutants. Test scores are often more reflective of test anxiety, previous classroom and schooling experiences, stereotype threat, and the prevalence of teachers teaching to the test than they are of actual abilities. Thus, although initial test scores may be an indication of the feasible improvement, the data revealed by test scores are too muddled with other information to make them reliable indicators of cadets' true abilities.

# References

[1] National Guard Youth Foundation. 2016. "About Youth ChalleNGe." National Guard Youth Foundation. Accessed Jan. 13, 2016. http://www.ngyf.org/about-challenge-2/.

[2] National Guard Youth ChalleNGe Program. 2015. "Core Components." Accessed Jan. 13, 2016. https://www.jointservicessupport.org/ngycp/core-components/EeqO-z8MgUK9NFRHRW8flQ.

[3] Wenger, Jennie W., and Jennifer R. Atkin. 2013. *Cognitive and Noncognitive Changes From Participation in National Guard Youth ChalleNGe.* CNA. DRM-2013-U-005201-Final.

[4] Sutton, Warren T., Lauren R. Malone, and Jennifer R. Atkin. 2014. *Cognitive and Noncognitive Improvements Among Cadets at the Washington Youth Academy (WYA).* CNA. DRM-2014-U-008846-Final.

[5] Wenger, Jennie W., and Cathleen M. McHugh. 2008. *Variation in Participants and Policies Across ChalleNGe Programs.* CNA. CRM D0017743.A2/Final.

[6] Aronson, Joshua. 2002. "Improving Academic Achievement: Impact of Psychological Factors on Education." In *Educational Psychology Series*. Edited by Gary D. Phye. San Diego, CA: Academic Press.

[7] Bandura, Albert. 1986. *Social Foundations of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

[8] Deming, David J. 2015. *The Growing Importance of Social Skills in the Labor Market*. National Bureau of Economic Research. NBER Working Paper 21473. Accessed Jan. 13, 2016. http://www.nber.org/papers/w21473.

[9] Levin, Henry M. 2012. "More Than Just Test Scores." *Prospects* 42: 269-284.

[10] Usher, Ellen L., and Frank Pajares. 2008. "Sources of Self-Efficacy in School: Critical Review of the Literature and Future Directions." *Review of Educational Research* 78 (4): 751-796.

[11] Heckman, James J., and Yona Rubinstein. 2001. "The Importance of Noncognitive Skills: Lessons From the GED Testing Program." *The American Economic Review* 91 (2): 145-149.

[12] Heckman, James J., Jingjing Hsse, and Yona Rubinstein. 2000. *The GED Is a "Mixed Signal": The Effect of Cognitive and Non-Cognitive Skills on Human*

*Capital and Labor Market Outcomes.* Unpublished University of Chicago Manuscript. Accessed Jan. 21, 2016. http://www.ssc.wisc.edu/econ/Durlauf/HHR.pdf.

[13] Heckman, James J., and Tim Kautz. 2014. "Fostering and Measuring Skills Interventions That Improve Character and Cognition." In *The Myth of Achievement Tests: The GED and the Role of Character in American Life.* Edited by James J. Heckman, John Eric Humphries, and Tim Kautz. Chicago: University of Chicago Press.

[14] Heckman, James, Rodrigo Pinto, and Peter Savelyev. 2013. "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes." *American Economic Review* 103 (6): 2052-2086.

[15] Heckman, James J., Seong Hyeok Moon, Rodrigo Pinto, Peter A. Savelyev, and Adam Yavitz. 2010. "The Rate of Return to the High/Scope Perry Preschool Program." *Journal of Public Economics* 94 (1-2): 114-128.

[16] Duckworth, Angela L., and Martin E. P. Seligman. 2005. "Self-Discipline Outdoes IQ in Predicting Academic Performance of Adolescents." *Psychological Science* 16 (12): 939-944.

[17] Bandura, Albert. 1977. "Self-Efficacy: Toward a Unifying Theory of Behavioral Change." *Psychological Review* 84 (2): 191-215.

[18] Bandura, Albert. 1997. "Self-Efficacy Concerns of Adulthood." In *Self-Efficacy: The Exercise of Control.* 184-198. New York: W. H. Freemand and Company.

[19] Heckman, James J. 2008. "Schools, Skills, and Synapses." *Economic Inquiry* 46 (3): 289-324.

[20] Cameron, Stephen V., and James J. Heckman. 1998. "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males." *The Journal of Political Economy* 106 (2): 262-333.

[21] Cunha, Flavio, and James J. Heckman. 2010. *Investing in Our Young People.* National Bureau of Economic Research. NBER Working Paper 16201.

[22] Garcia, Emma. 2014. *The Need To Address Noncognitive Skills in the Education Policy Agenda.* Economic Policy Institute. Briefing Paper #386.

[23] Anger, Silke. 2011. *The Intergenerational Transmission of Cognitive and Non-Cognitive Skills During Adolescence and Young Adulthood.* The German Socio-Economic Panel Study (SOEP) at DIW Berlin. Accessed Nov. 30, 2015. http://www.iza.org/conference_files/TAM2012/anger_s975.pdf.

[24] Coleman, Margo, and Thomas DeLeire. 2003. "An Economic Model of Locus of Control and the Human Capital Decision." *The Journal of Human Resources* 38 (3): 701-721.

[25] Pajares, Frank. 2005. "Self-Efficacy During Childhood and Adolescence: Implications for Parents and Teachers." In *Self-Efficacy Beliefs of Adolescents*. Edited by Frank Pajares and Tim Urdan. 339-368. Greenwich, CT: IAP Publishing.

[26] Stipek, Deborah J., and John R. Weisz. 1981. "Perceived Personal Control and Academic Achievement." *Review of Educational Research* 51 (1): 101-137.

[27] Rotter, Julian B. 1975. "Some Problems and Misconceptions Related to the Construct of Internal Versus External Control of Reinforcement." *Journal of Consulting and Clinical Psychology* 43 (1): 56-67.

[28] Cairns, Ed, Liz McWhirter, Ursulla Duffy, and Robert Barry. 1990. "The Stability of Self-Concept in Late Adolescence: Gender and Situational Effects " *Personality and Individual Differences* 11 (9): 937-944.

[29] Hendricks, Leo E., and Teresa A. Montgomery. 1984. "Educational Achievement and Locus of Control Among Black Adolescent Fathers." *The Journal of Negro Education* 53 (2): 182-188.

[30] National Education Association. 2015. *Backgrounder: Indicators of Future Success: GPA and Noncognitive Skills.* Accessed Oct. 20, 2015. http://www.nea.org/assets/docs/Indicators_of_Success-BGH_ac5-final.pdf.

[31] Uguroglu, Margaret E., and Herbert J. Walberg. 1979. "Motivation and Achievement: A Quantitative Synthesis." *American Educational Research Journal* 16 (4): 375-389.

[32] Usher, Ellen L. 2009. "Sources of Middle School Students' Self-Efficacy in Mathematics: A Qualitative Investigation." *American Educational Research Journal* 46 (1): 275-314.

[33] Frome, Pamela M., and Jacquelynne S. Eccles. 1998. "Parents' Influence on Children's Achievement-Related Perceptions." *Journal of Personality and Social Psychology* 74 (2): 435-452.

[34] Eccles, Jacquelynne, Carol Midgley, and Terry F. Alder. 1984. "Grade-Related Changes in the School Environment: Effects on Achievement Motivation." In *Advances in Motivation and Achievement: The Development of Achievement Motivation*. Edited by John G. Nicholls. 283-331. Greenwich, CT: Jai Press, Inc.

[35] Schunk, Dale H., and Frank Pajares. 2002. "The Development of Academic Self Efficacy." In *Development of Achievement Motivation*. Edited by A. Wigfield, and J. Eccles. 16-29. San Diego: Academic Press.

[36] Chubb, Nancy H., and Carl I. Fertman. 1997. "Adolescent Self-Esteem and Locus of Control: A Longitudinal Study of Gender and Age Differences." *Adolescence* 32 (125): 113-117.

[37] Pajares, Frank. 1996. "Self-Efficacy Beliefs in Academic Settings." *Review of Educational Research* 66 (4): 543-578.

[38] Heckman, James J. 2000. "Policies To Foster Human Capital." *Research in Economics* 54: 3-56.

[39] Kautz, Tim, James J. Heckman, Ron Diris, Bas ter Weel, and Lex Borghans. 2014. *Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills To Promote Lifetime Success.* IZA Discussion Paper No. 8696.

[40] Cunha, Flavio, James J. Heckman, Lance Lochner, and Dimitriy V. Masterov. 2006. "Interpreting the Evidence on Life Cycle Skill Formation." In *Handbook of the Economics of Education, Volume 1*. Edited by Eric A. Hanushek, and Finish Welch. 697-812. Amsterdam: Elsevier North-Holland.

[41] Smith, Elizabeth S. 1999. "The Effects of Investments in the Social Capital of Youth on Political and Civic Behavior in Young Adulthood: A Longitudinal Analysis." *Political Psychology* 20 (3): 553-580.

[42] Knudsen, Eric I., James J. Heckman, Judy L. Cameron, and Jack P. Shonkoff. 2006. "Economic, Neurobiological, and Behavioral Perspectives on Building America's Future Workforce." *Proceedings of the National Academy of Sciences of the United States of America* 103 (27): 10155-10162.

[43] Mirowsky, John. 1995. "Age and the Sense of Control." *Social Psychology Quarterly* 58 (1): 31-43.

[44] Roberts, Brent W., Kate E. Walton, and Wolfgang Viechtbauer. 2006. "Patterns of Mean-Level Change in Personality Traits Across the Life Course: A Meta-Analysis of Longitudinal Studies." *Psychological Bulletin* 132 (1): 1-25.

[45] Wigfield, Allan, Jacquelynne S. Eccles, and Paul R. Pintrich. 1996. "Development Between the Ages of 11 and 25." In *Handbook of Educational Psychology*. 148-185. New York: Macmillan Library Reference.

[46] Farrington, C. A., M. Roderick, E. Allensworth, J. Nagaoka, T. S. Keyes, D. W. Johnson, and N. O. Beechum. 2012. *Teaching Adolescents To Become Learners.* University of Chicago Consortium on Chicago School Research.

[47] Cunha, Flavio, James J. Heckman, and Susanne M. Schennach. 2010. "Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Econometrica* 78 (3): 883-931.

[48] Tough, Paul. 2012. *How Children Succeed*. New York: Houghton Mifflin Harcourt.

[49] Almlund, Mathilde, Angela Lee Duckworth, James J. Heckman, and Tim D. Kautz. 2011. *Personality Psychology and Economics.* National Bureau of Economic Research. NBER Working Paper 16822. Accessed Nov. 30, 2015. http://www.nber.org/papers/w16822.

[50] Armstrong, William B. 2000. "The Association Among Student Success in Courses, Placement Test Scores, Student Background Data, and Instructor

Grading Practices." *Community College Journal of Research and Practice* 24 (8): 681-695.

[51] Scott-Clayton, Judith. 2012. *Do High-Stakes Placement Exams Predict College Success?* CCRC Working Paper No. 41.

[52] Haladyna, Thomas M., Susan Bobbit Nolen, and Nancy S. Haas. 1991. "Raising Standardized Achievement Test Scores and the Origins of Test Score Pollution." *Educational Researcher* 20 (5): 2-7.

[53] Good, Catherine, Aneeta Rattan, and Carol S. Dweck. 2012. "Why Do Women Opt Out? Sense of Belonging and Women's Representation in Mathematics." *Journal of Personality and Social Psychology* 102 (4): 700-717.

[54] Schwartz, Tony. 1999. "The Test Under Stress." *New York Times Magazine.* Jan. 10, 1999.

[55] Piccone, Jason E. 2006. "Administering the Test of Adult Basic Education at Intake: A Biased Marker of Offender Ability." *The Journal of Correctional Education* 57 (3): 239-248.

[56] Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2011. *The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood.* National Bureau of Economic Research. NBER Working Paper 17699. Accessed Dec. 2, 2015. http://www.nber.org/papers/w17699.

[57] Goldhaber, Dan D., Dominic J. Brewer, and Deborah J. Anderson. 1999. "A Three-way Error Components Analysis of Educational Productivity." *Education Economics* 7 (3): 199-208.

[58] Steele, Claude M., Steven J. Spencer, and Joshua Aronson. 2002. "Contending With Group Image: The Psychology of Stereotype and Social Identity Threat." *Advances in Experimental Psychology* 34: 379-440.

[59] Appel, Markus, Silvana Weber, and Nicole Kronberger. 2015. "The Influence of Stereotype Threat on Immigrants: Review and Meta-analysis." *Frontiers in Psychology* 6: 1-15.

[60] Steele, Claude M., and Joshua Aronson. 1995. "Stereotype Threat and the Intellectual Test Performance of African Americans." *Journal of Personality and Social Psychology* 69 (5): 797-811.

[61] Good, Catherine, Joshua Aronson, and Michael Inzlicht. 2003. "Improving Adolescents' Standardized Test Performance: An Intervention To Reduce the Effects of Stereotype Threat." *Applied Developmental Psychology* 24: 645-662.

[62] Steele, Claude M. 1992. "Race and the Schooling of Black Americans." *The Atlantic Monthly.* Apr. 1992.

[63] Katz, Irwin, S. Oliver Roberts, and James M. Robinson. 1965. "Effects of Task Difficulty, Race of Administrator, and Instructions on Digit-Symbol

Performance of Negroes." *Journal of Personality and Social Psychology* 2 (1): 53-59.

[64] Mehrens, William A., and John Kaminski. 1989. "Methods for Improving Standardized Test Scores: Fruitful, Fruitless, or Fraudulent?" *Educational Measurement: Issues and Practice* 8 (1): 14-22.

[65] Shephard, Lorrie A. 1990. "Inflated Test Score Gains: Is the Problem Old Norms or Teaching to the Test?" *Educational Measures, Issues, and Practice* 9 (3): 15-21.

[66] Koretz, Daniel M. 2002. "Limitations in the Use of Achievement Tests as Measures of Educators' Productivity." *The Journal of Human Resources* 37 (4): 752-777.

[67] Cunha, F., and J. J. Heckman. 2008. "Formulating, Identifying, and Estimating the Technology of Cognitive and Non-Cognitive Skill Formation." *The Journal of Human Resources* 42 (4): 738-782.

[68] Haertel, Edward H. 2013. *Reliability and Validity of Inferences About Teachers Based on Student Test Scores.* Educational Testing Service. Accessed Jan. 26, 2016. https://www.ets.org/s/pdf/23497_Angoff%20Report-web.pdf.

[69] Wilson, T. D., and P. W. Linville. 1985. "Improving the Performance of College Freshmem With Attributional Techniques." *Journal of Personality and Social Psychology* 49: 287-293.

[70] Aronson, Joshua, Carrie B. Fried, and Catherine Good. 2002. "Reducing the Effects of Stereotype Threat on African American College Students by Shaping Theories of Intelligence." *Journal of Experimental Social Psychology* 38: 113-125.

This page intentionally left blank.

# The CNA Corporation

This report was written by CNA Corporation's Resource Analysis Division (RAD).

RAD provides analytical services—through empirical research, modeling, and simulation—to help develop, evaluate, and implement policies, practices, and programs that make people, budgets, and assets more effective and efficient. Major areas of research include health research and policy; energy and environment; manpower management; acquisition and cost; infrastructure; and military readiness.

**DRM-2016-U-012883-Final**




CNA Corporation is a not-for-profit research organization
that serves the public interest by providing
in-depth analysis and result-oriented solutions
to help government leaders choose
the best course of action
in setting policy and managing operations.

*Nobody gets closer—*
*to the people, to the data, to the problem.*

CNA
ANALYSIS & SOLUTIONS

www.cna.org • 703-824-2000

3003 Washington Boulevard, Arlington, VA 22201